This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2851451, IEEE Access

State of the Art: Eye-Tracking Studies in Medical Imaging

Lucie Lévêque, Hilde Bosmans, Lesley Cockmartin and Hantao Liu, *Member, IEEE*

***Abstract*—Eye-tracking – the process of measuring where people look in a visual field – has been widely used to study how humans process visual information. In medical imaging, eye-tracking has become a popular technique in many applications to reveal how visual search and recognition tasks are performed, providing information that can improve human performance. In this paper, we present a comprehensive review of eye-tracking studies conducted with medical images and videos for diverse research purposes, including identification of degree of expertise, development of training, and understanding and modelling of visual search patterns. In addition, we present our recent eye-tracking study that involves a large number of screening mammograms viewed by experienced breast radiologists. Based on the eye-tracking data, we evaluate the plausibility of predicting visual attention by computational models.**

***Index Terms*—Medical imaging, visual attention, image quality, eye-tracking, saliency**

I. Introduction

EYE-TRACKING is a widely used methodology which enables recording of eye positions and movements of a human subject for further interpretation and applications. In fact, eye movements allow a deeper insight into human attention, up to revealing their needs and emotional states for instance [1]. The phenomenon of human visual attention has been studied for over a century, with the objective to understand how human brain continuously minimises the overloading amount of input into a manageable flow of information. Significant findings were established in the literature that visual attention is essentially driven by two general attentional processes, i.e., *bottom-up* and *top-down* [2]. Bottom-up aspects are based on the characteristics of the visual scene, making it stimulus driven. Regions of interest which attract attention in a bottom-up way must be sufficiently distinctive with respect to surrounding features [3]. On the other hand, top-down attention is driven by factors such as knowledge, expectation and experience. Eye-tracking, and more particularly the measurement of the point of gaze, has emerged as the key means of studying visual attention. Origins of eye-tracking date back to 1879 when Louis Emile-Javal, a French ophthalmologist, noticed based on naked-eye observations that readers' eyes make quick movements (i.e., saccades) mixed with short pauses (i.e., fixations) while reading. The first eye-tracker, which was an intrusive device, was built in 1908 by Edmund Huey. The first non-intrusive recordings of eye movements were conducted by Guy Thomas Buswell, an educational psychologist, in 1937 [4]. During the 1970s and 1980s, video-based eye-trackers were invented to enable less intrusive and more accurate eye-tracking practice. It is nowadays used in a wide range of applications, including cognitive psychology, marketing research, usability engineering, human computer interaction, and medical image quality [5]. An eye-tracking study usually involves the participation of a certain number of human subjects, the recording of their eye movements using a sophisticated eye-tracker, and the agglomerated analysis of their fixation/gaze patterns.

In recent years, there has been a growing interest in the use of eye-tracking technology in medical imaging. Medical images are not self-explanatory and thus need to be viewed and interpreted by medical professionals [6]. However, the interpretation task is not always easy and even competent clinicians can make errors mainly due to the limitations of the human eye-brain system. Estimates indicate that, in some areas of radiology, the miss (i.e., false negative) rate may be as high as 30%, with an equally high false-positive rate [7]. Errors can have significant impact on patient care and it is therefore important to understand how humans understand medical images so that we can improve their diagnostic performance [8]-[9]. In radiology for example, eye-tracking methodologies have been widely used to study how visual search and recognition tasks are performed, providing information that can improve speed and accuracy of radiological reading. Generally, in a typical eye-tracking study, a target stimulus is presented to a sample of image readers while their eye movements are recorded by an eye-tracker. The resulting eye-tracking data are then statistically analysed to provide evidence of the subjects' visual behaviour. This information can be subsequently used to assess the image quality of diagnostic imaging systems and to improve the task performance of medical professionals. Also, it would be highly beneficial for image readers to have a tool that can automatically and accurately predict where experts look in images. This can be used as an automated perceptual feedback system to enhance their diagnostic performance.

L. Lévêque and H. Liu are with the School of Computer Science and Informatics, Cardiff University, Cardiff, CF24 3AA UK (e-mail: levequel@cardiff.ac.uk and liuh35@cardiff.ac.uk, respectively).

H. Bosmans and L. Cockmartin are with the Department of Radiology, University Hospitals KU Leuven, Leuven, 3000 Belgium (e-mail: hilde.bosmans@uzleuven.be and lesley.cockmartin@uzleuven.be, respectively).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2851451, IEEE Access

In this paper, we present a comprehensive literature review that focuses on eye-tracking studies in medical imaging. Most of the existing surveys target a specific imaging modality or a specific clinical disease, whereas this survey contains diverse fields and applications in medical imaging. We discuss the existing eye-tracking studies: the visual search patterns will be reviewed in section II; the study of the influence of expertise will be summarised in section III; and the work relating to the impact of training on viewing behaviour will be surveyed in section IV. Furthermore, we present our recent eye-tracking study of screening mammograms in section V, where we discuss the importance and challenges of automatically predicting eye movements and aim to evaluate to what extent a computational model can predict the gaze of experts, as this was found of potentially crucial importance for large scale practical applications of improved teaching.

II. Visual Search Patterns in Medical Imaging

It is important to identify visual search patterns that are associated with high perceptual performance, and consequently to determine optimal visual search strategies. A summary of the eye-tracking studies discussed in this section is detailed in Table I.

TABLE I
OVERVIEW OF THE EYE-TRACKING STUDIES INVESTIGATING THE VISUAL SEARCH PATTERNS OF MEDICAL PROFESSIONALS

Source	Carmody et al. [10]	Beard et al. [11]	Suwa et al. [12]	Kundel et al. [13]	Voisin et al. [14]	Almansa et al. [15]	Drew et al. [16]
Modality	10 chest X-ray slides (4 normal, 6 abnormal)	3 cases of CT scans (single chest and multiple abdominal)	20 CT images (10 normal, 10 with pathologic lesions)	Mammograms	40 mammograms (20 malignant, 20 benign)	3 HD segments of colonoscopy	5 3D chest CTs
Task	Press a key when finding nodule	Read CTs and report interpretation	Discriminate normal and pathologic images, indicate pathologic site	Detect lesions	Give the probability of malignancy of the mass	Detect adenoma	Detect as many nodules as possible in three minutes
Partici-pants	4 radiologists	4 radiologists, 1 radiology resident	8 dentists (dental radiologists, oral surgeons)	3 expert mammogra-phers, 3 fellows, 3 residents	2 expert mammogra-phers, 4 radiology students	11 endoscopists	24 radiologists
Eye-tracker	Set of glasses with a corneal reflection technique	Eye Mark Recorder model V	Eye-tracking system model 504	ASL	Mirametrix S2	ASL Mobile Eye	EyeLink 1000
Gaze metrics	Eye movements and fixations	Not specified	Time to discriminate, fixation point count, travel distance between fixations, average time per fixation	Not specified	Number of fixations, duration of fixations, fixation/saccade ratio, saccade length	Total gaze time, number of times gaze is directed to each segment of the screen, mean duration	Saccadic amplitude, mean number of quadrant fixation clusters
Main findings	Radiologists use a comparative scanning strategy to differentiate nodules from anatomical structures	Sequential overview of the images followed by a detailed examination of the clusters	Tendency to scan for ROI on normal images, tendency for interpretation by concentrating on suspect regions for pathologic images	57% of cancer locations are fixated within the first second of screening	Diagnostic error is correlated with duration and number of fixations (longer review time correlates with higher chance of error)	Positive correlation between detection rate and central gaze pattern, experienced endoscopists spend less time on the centre	Two strategies: scanners (search through a given slice, strong organisation) and drillers (limit search to a quadrant, found more lung nodules)

In 1981, Carmody et al. [10] published one of the first eye-tracking studies where visual search was investigated by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2851451, IEEE Access

means of eye-position recording techniques. They studied the detection of lung nodules in chest X-ray films. Four radiologists participated in the experiment, where they were asked to search for nodules in ten chest films. Their eye movements were recorded using special glasses based on corneal reflection technique. The subjects were instructed to press a key when they found a nodule in the X-rays. The eye-tracking data, i.e., visual dwell times were used to analyse visual search behaviour. It was found that false negative (omission) errors were impacted by both the visibility of the nodule and the scanning strategies used by the radiologist.

A decade later, Beard et al. [11] conducted an eye-tracking study using an Eye Mark Recorder (model V) to understand visual scan patterns developed by radiologists when interpreting both single chest and multiple abdominal computerised tomography (CT) scans. Four radiologists and one radiology resident participated in the first part of the study where single CT scans were tested. Their task was to read and interpret three patient cases, each of which contains 30 to 40 image slices. Radiologist scan patterns were rendered manually from the tape records; and a systematic sequential visual scan pattern was found. The second part of the study was to assess how images were cross compared, using multiple CT scans. The radiologists had to view three patient folders each containing more than one CT scan with the number of films exceeding the available viewing space. Eye-tracking data showed that the radiologists used a similar approach of reading single CT scans, i.e., a systematic sequential visual scan, however, they also developed a comparison method.

Suwa et al. [12] also carried out a study with CT images but in the field of dentistry. They recruited eight dentists, and each was shown ten normal and ten pathologic CT images. Eye movements of the dentists were recorded with an eye-tracking system (model 504) when interpreting the images. Six parameters were extracted from the eye-tracking data, namely time to determinate whether the image is normal or pathologic, fixation point count, distance between fixations, time spent on each fixation, total gaze fixation time, and minimum gaze fixation time. Based on these parameters, the gaze patterns of the dentists were investigated. In terms of the difference in gaze patterns between normal and pathologic images, it was found that when viewing a normal image, the subjects tended to move sequentially (as noticed by Beard et al. [11]), whereas, when viewing a pathologic image, the tendency was to focus on suspected regions. Moreover, they found that both the travel distance between fixations and the minimum gaze fixation time were longer for pathologic images than normal images. The total gaze fixation time, which is shorter for normal images, significantly contributed to determine whether an image was normal or pathologic.

Eye-tracking studies were also conducted in other areas of specialties, such as mammography. Kundel et al. [13] gathered eye-tracking data collected independently at three institutions with an ASL (Applied Science Laboratories) eye-tracking device, where experienced mammographers, mammography fellows, and radiology residents searched for cancers in mammograms, both on craniocaudal and mediolateral oblique views. They found that 57% of cancer locations were fixated within the first second of viewing. They concluded that the initial detection occurs before visual scanning and that the development of expertise may consist of a shift from scan-look-detect to look-detect-scan mechanism.

Voisin et al. [14] also worked on mammogram images. They investigated the association between gaze patterns and diagnostic performance for lesion detection in mammograms. They recorded the eye movements of six radiologists while evaluating the likelihood of malignancy of forty mammographic masses, using a Mirametrix S2 eye-tracker. By assessing various quantitative metrics derived from the eye-tracking data, such as the fixation duration, number of fixations, and fixation/saccade ratio, they showed that these gaze metrics were highly correlated with radiologists' diagnostic errors.

Almansa et al. [15] investigated the relationship between gaze patterns captured with an ASL mobile eye-tracking device and adenoma detection rate in colonoscopy videos. Eleven endoscopists participated in a study in which they were asked to watch three high-definition video clips from three normal colonoscopies. Diverse forms of information were gathered from the eye-tracking data, including the total gaze time, number of fixations, and mean duration of fixations. The results showed that the adenoma detection rate was significantly correlated with the central gaze time, i.e., the time spent on the centre of the screen. It was found that the participants who detected the highest number of adenomas showed a tendency to focus on the centre of the screen, whereas the participants who detected less lesions moved their eyes more broadly around.

Drew and al. [16] worked on 3D CT images. Twenty-four radiologists were recruited to search for lung nodules in chest CT scans. Five cases were used, and there were fifty-two nodules in total. The radiologists were asked to find as many nodules as possible in three minutes. Based on the eye-tracking data collected using an EyeLink1000 eye-tracking device, Drew et al. divided the radiologists into two groups depending on their reading strategies: the "scanners" and the "drillers". The scanners usually search throughout a slice in depth before moving to a new depth, whereas the drillers limit their search to a part of the lung while scrolling through slices in depth. In general, drillers found more nodules than scanners.

III. Influence of Experience and Expertise in Medical Imaging

To improve the diagnostic performance of less experienced readers, it is necessary to understand how they perceive medical images and then to compare their viewing behaviour with expert readers. Existing eye-tracking studies that compare viewing behaviour of experts and novices can be divided into two categories: studies on medical diagnosis (see Table II) and studies on surgery (see Table III). We will discuss each category in detail below.

A. Diagnosis

Table II summarises the studies that compare experts and novices when rendering diagnoses based on diverse modalities of medical imaging, including, but not limited to, computed tomography (CT), magnetic resonance imaging (MRI) and radiographs.

Nodine et al. [17] carried out an eye-tracking experiment where the participants (i.e., three mammographers and six radiology trainees) were asked to view 40 mammogram cases and decide whether they were "normal" or "abnormal". Their eye movements were recorded using an ASL4000 eye-head tracker. Experimental results showed there was no significant difference in terms of the decision time between experts and trainees, however, the performance of mammographers was always higher than trainees. The eye-fixation patterns of trainees were compared to that of experienced mammographers; and the results indicated that trainees did not spend enough time on the lesions.

Similar findings were obtained in the study of Tourassi et al. [18], where three breast imaging radiologists and three residents were asked to view 20 screening mammograms for a specific task of mass detection while wearing a H6 head-mounted eye-tracker. In consistence with the study of Nodine et al. [17], the residents' detection accuracy was on average lower than the experts. The recall rate of residents and expert radiologists was nonetheless the same on average. The results also showed that radiologists have a more complex gaze behaviour than residents.

TABLE II
OVERVIEW OF THE EYE-TRACKING STUDIES INVESTIGATING THE IMPACT OF EXPERIENCE IN RADIOLOGY

Source	Nodine et al. [17]	Tourassi et al. [18]	Cooper et al. [19]	Matsumoto et al. [20]	Bertram et al. [21]	Bertram et al. [22]
Modality	40 mammography test cases (20 with lesions, 20 without)	20 screen-films mammograms	CT and MRI brain images with strokes	Brain CT images after cerebro-vascular accident	Abdominal CTs	26 abdominal CTs
Task	Make a decision: normal or abnormal	Give a diagnosis	Detect abnormalities	Search for lesions and give a diagnosis	Detection of abnormalities, detection of lymph nodes	Detect lesions
Partici-pants	3 radiographers, 6 radiography trainees	3 radiologists, 3 radiography residents	3 experienced readers, 1 trainee, 4 novices	12 neurologists and 12 control subjects (nurses, psychologists, medical students)	7 radiologists, 9 radiographers, 22 students	12 specialists, 14 advanced residents, 15 early residents
Eye-tracker	ASL 4000 SU	H6 head-mounted	Tobii x50	EyeLink 100	EyeLink 1000	EyeLink 1000
Gaze metrics	Not specified	Dwells, initials, returns	Not specified	Not specified	Fixation duration, saccade lengths	Fixation duration, saccade lengths
Main findings	Mammogra-phers' performance is higher than trainees'	Residents' detection accuracy lower than experts, radiologists' behaviour predicted by local content, cognitive behaviour predicted by observing gaze	Experts employed a strategy by spending more time on the AOI than novices and trainee	Neurologists and controls differ in behaviour: both look at high-saliency areas but neurologists gaze more at clinically important areas	Experts performed better than semi-experts and novices	Novices detected less low contrast lesions, experts have shorter saccades length when lesions

Source	Mallett et al. [23]	Manning et al. [24]	Leong et al. [25]	Vaidyanathan et al. [26]	Turgeon et al. [27]
Modality	28 endoluminal 3D CT colonography	120 digitised post anterior chest	33 skeletal radiographs	34 dermatological images	20 dental panoramic images

	videos	images	(shoulder, hand, knee)		
Task	Find and click on polyps		Search for the fracture(s) and press a button	For experts, describe as if teaching to make a diagnosis; for trainees, ask questions	Free viewing
Partici-pants	27 experienced radiologists, 38 inexperienced radiologists	8 experienced radiologists, 5 radiographers, 8 novice radiography students	25 observers: radiologists, ortho surgeons, specialist registrars, ortho and A&E senior house officers	12 dermatology experts, 12 undergraduate novices	30 dental students, 15 facial and maxillofacial radiologists (OMRs)
Eye-tracker	Tobii x50 and Tobii x120	ASL 504 remote	Tobii 1750	SMI 50 Hz	SMI RED-m
Gaze metrics	Time to first pursuit, assessment pursuit time, assessment pursuit rate	Fixations per image, mean saccadic amplitudes per image, coverage of image area, total duration of film scrutiny	Not specified	Median fixation duration, saccade amplitudes	Total distance tracked, time to first fixation, total duration and number of fixations for the ROI
Main findings	Experts have a higher rate of identification but a similar percentage of pursuit	Radiologists and radiographers after training are better at the task than novices	Experts showed higher number of true positives, with less dwell time on the fracture site	Experts are able to weigh a region's importance after a brief fixation	OMRs required less time for images with pathosis, they made fewer fixations and saccades, and took less time before first fixation at the ROI

There are few studies that focus on CT images, such as Cooper et al. [19], Matsumoto et al. [20], Bertram et al. [21]-[22], and Mallett et al. [23]. Cooper et al. [19] investigated visual search behaviour on stroke images with three experienced readers, one trainee and four novices. The participants were asked to rate eight clinical cases on a five-point Likert scale, depending on the presence or absence of abnormality and their degree of confidence. The results showed there was a significant difference in diagnostic accuracy between novices and experts; the experts performed better than the novices. The recorded eye-tracking data were used to reveal the reasoning behind the observed difference between novices and experts. In the case of an acute stroke, the trainee reader noticed the region of interest with the 34^{th} fixation whereas the experts fixated in with their first fixation. For a chronic stroke case, the novices only spent a short time looking at the affected area, and the experts concentrated on the affected tissue from the first fixation. Matsumoto et al. [20] also studied stroke cases two years later, with twelve neurologists and twelve control subjects consisting of nurses, medical technologists, psychologists and medical students. The findings were that both neurologists and control subjects gazed at visually salient areas in the images, however, only the neurologists gazed at visually low-salient areas with clinical importance. Bertram et al. [21]-[22] applied the approach of the two studies mentioned above to abdominal CT images. In their first study [21], they compared the eye movements of seven radiologists, nine radiographers and twenty-two psychology students when watching abdominal CT scans. The participants had to perform an easy task, i.e., the detection of visually salient abnormalities, and a difficult task, i.e., the detection of enlarged lymph nodes. Results showed that for the difficult task, experts performed better than semi-experts and naïve participants; however, there was no difference in detection performance between semi-experts and novices. For the easy task, experts and semi-experts performed better than naïve participants. In the second study [22], Bertram et al. investigated markers of visual expertise using 26 abdominal CT images. An eye-tracking experiment was conducted with twelve specialists, fifteen advanced residents and fifteen early residents when performing a detection task. Similar to their first study, they found that the detection rate of specialists was higher than that of residents, and that advanced residents detected more lesions than early residents. On average, eye-tracking data showed that specialists reacted to the presence of lesions using long fixation durations and short saccades. Finally, Mallett et al. [23] focused their study on 23 3D CT colonography videos, which were interpreted by twenty-seven experienced and thirty-eight inexperienced radiologists. Experimental results showed that experienced readers had a higher rate of polyp identification than inexperienced readers, but there was no difference between the two groups in terms of percentage of pursuits and total assessment period. Eye-tracking data revealed that readers examined polyps by multiple pursuits, which means they recognised the importance of the lesions. There was no difference in the rate of scanning errors between experienced and inexperienced readers.

The scope of eye-tracking studies was broadened by Manning et al. [24], Leong et al. [25], Vaidyanathan et al. [26], and Turgeon et al. [27], for radiographs, chest images,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2851451, IEEE Access

dermatological images, and panoramic images, respectively. Manning et al. [24] analysed the gaze behaviour of eight experienced radiologists, five experienced radiographers (before and after training) and eight undergraduate radiography students when detecting nodules, with an ASL504 remote eye-tracking device. They showed that the radiologists and radiographers after training were better at performing the task than the novices, and that the novices and radiographers before training made more fixations per film. In the study of Leong et al. [25], they recruited twenty-five observers with different specialisation who had to examine 33 skeletal radiographs and identify the fractures. Their eye movements were recoded using a Tobii 1750 eye-tracker. The results showed that there was no significant difference between the groups in the time spent on evaluating the radiographs. However, the experts had a higher number of true positives. Vaidyanathan et al. [26] compared the eye movements of twenty-two dermatology experts and twelve undergraduate novices when viewing 34 dermatological images. Their main finding is that experts can weigh a region's importance after a brief fixation, whereas novices need multiple re-fixations. Moreover, they found that the median fixation duration and saccade amplitude are significantly higher for experts than for novices. Finally, in a more recent study, Turgeon et al. [27] used 20 dental panoramic images to assess the influence of experience on eye movements with a SMI RED-m device. They asked fifteen oral and maxillofacial radiologists and thirty dental students to view freely the images, while their gaze movements were recorded. They found that all participants spent more time on normal images than abnormal images. Radiologists needed less time before making their first fixation on the region of interest, and they made fewer fixations than the students on images of pathoses.

To summarise, the results from different eye-tracking studies showed that experts and novices have different gaze behaviours when making diagnoses based on medical images. Novices should be trained to get the expert level characterised by a particular gaze behaviour.

B. *Surgery*

Table III summarises the studies that compare experts and novices when evaluating surgical images or videos.

Law et al. [28] are the first researchers who investigated the gaze behaviour between experts and non-experts for laparoscopic surgery in 2004. They had the hypothesis that there would be distinctive characteristics in gaze between the two groups of subjects. In present survey, we will compare the differences to what has been observed in radiology. Law et al. conducted an eye-tracking experiment with five expert surgeons and five students, where the subjects had to perform a virtual task: they had to touch a small target using a virtual laparoscopic tool, as quickly as possible and without error if possible, for 2 blocks of 5 trials each. Eye-tracking data were collected using an ASL 504 remote eye-tracking device. The results showed that the experts performed significantly better than non-expert participants, both in time and precision. In terms of visual behaviour, the novices spent more time looking at the tool than the experts.

Kocak et al. [29] then recorded the eye movements of eight novices, eight intermediates and eight experts in surgery with a Cyclops Eye Trak saccadometer when performing three basic laparoscopic tasks, i.e., loops, rope and beans. The results showed that the degree of experience affected the fixations and saccades. The average saccadic rate was significantly higher for the novices than the experts. Furthermore, the duration of fixations was higher for the expert group than the intermediate group and the novice group.

In 2010, Ahmidi et al. [30] published their eye-tracking study on laparoscopic surgery. They recruited five expert surgeons and six novices who had to find a given anatomy in the sinus cavity and touch it using an endoscope. The work showed that the surgeons' gaze data included skill related structures, which were, however, not found for novices. They also presented an objective method to assess the expertise level of surgeons using the Hidden Markov Model.

At the same time, Richstone et al. [31] published their study. Twenty-one surgeons participated in a simulated and live surgery where they had to achieve different tasks of varying degrees of difficulty. Their eye movements were recoded using an EyeLink II eye-tracker. Quantitative metrics related to eye movements, namely blink rate, fixation rate, pupil metric and vergence were evaluated. The work demonstrated that, both for the simulation study and live surgery, eye metrics can make a distinction between non-expert and expert surgeons in a reliable way.

Finally, Khan et al. [32] studied eye movements of surgeons when performing a surgical task and later on when watching the operative video, as well as the gaze of surgical residents. Two expert surgeons and twenty novices were recruited for the eye-tracking study using a Tobii X50 device. Sixteen laparoscopic cholecystectomy cases were used. The results showed that there was a 55% overlap for the expert surgeons between "doing" and "self-watching", and only 43.8% for the junior residents. The difference between the two groups is statistically significant.

All the eye-tracking studies available in the literature focus on laparoscopic surgery, which is a type of minimally invasive surgery. This practice is of benefit to patients due to the reduction of the incisions and of the recovery time. In general, in terms of the impact of expertise on gaze behaviour, the findings are similar to that of radiology studies.

IV. Impact of Training on Viewing Behaviour in Medical Imaging

In the previous section, we have discussed the differences in viewing behaviour less experienced readers and more experienced readers. The following question is how training plays a role in changing behaviour. Table IV summarises the eye-tracking studies that assess the impact of training on the viewing behaviour of medical professionals.

TABLE III
Overview of the eye-tracking studies investigating the impact of experience in Surgery

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2851451, IEEE Access

Source	Law et al. [28]	Kocak et al. [29]	Ahmidi et al. [30]	Richstone et al. [31]	Khan et al. [32]
Modality	Laparoscopic surgery	Laparoscopic surgery	Endoscopic sinus surgery	Laparoscopic surgery (simulation and live)	16 cases of laparoscopic chlolecystectomy
Task	Touch a virtual target cube using a laparoscopic instrument	3 basic laparoscopic tasks: loops, rope, beans	Find and touch a given anatomy	Different levels (easy, moderate, difficult)	Watching freely
Partici-pants	5 experts, 5 novices	8 experts, 8 intermediates, 8 novices	5 experts, 6 novices	11 surgeons for the live, 10 surgeons for the simulation	2 expert surgeons, 20 novices
Eye-tracker	ASL 504 remote	Cyclops Eye Trak saccadometer	Not specified	EyeLink II	Tobii X50
Gaze metrics	Not specified	Saccades, saccadic amplitude, peak velocity per saccade, duration of each saccade, duration of gaze fixation	Not specified	Blink rate, fixation rate, pupil metric, vergence	Not specified
Main findings	Experts completed the task more quickly (but no difference in accuracy), and showed more target-based behavior (novices: tool based)	Saccadic rate lower for experts than novices (intermediate in between), peak velocity higher for experts than for novices (intermediate in between)	Surgeon eye-gaze data include structure for eye level recognition	Fixation rate is higher for experts than trainees (both for simulation and live)	Experts showed greater overlap with the reference subjects (recorded videos) than novices

As we discussed in the previous section, expert surgeons tend to focus on their task whereas novices follow the tool during laparoscopic surgery. Wilson et al. [33] developed further research to study the effect of training on gaze behaviour in laparoscopic surgery with an ASL mobile eye-tracking device. Thirty medical trainees who had received no laparoscopic training participated in the experiments. They were divided into three equal groups; and each group received a different training program, i.e., gaze training, movement training, or discovery training. The first group was shown a video of an expert's eye movements when performing a coordination task. The second group was shown the same video but without the gaze cursor. Finally, the third group was given no video or instructions but was allowed to examine their own performance. Before training, statistical analyses showed there was no significant difference between the three groups in terms of completion time. After training, the results showed that the gaze group was significantly faster than the movement group and the discovery group. Furthermore, the gaze group spent significantly more time than the other two groups using target locking fixations, i.e., fixations spent on the target ball and not on the tool. It is suggested that the neural mechanisms in charge of goal-directed movements benefit from the foveated target [34].

Vine et al. [35] conducted a similar study to assess the impact of gaze training in laparoscopic surgery; however, in contrast to the study of Wilson et al. [33], the participants were not made aware of the objective of the training. Twenty-seven participants who had not received any laparoscopic training were involved in the study. They were assigned to a gaze training group or to a discovery learning group. Each participant had to complete a task, i.e., to move foam balls into a cup using a single instrument. The first group was shown a surgery training template to passively adopt experts' gaze patterns, whereas the second group did not use the template. There was no significant difference between the two groups before training. After training, statistical analyses revealed a significant difference between the two groups in terms of completion time and accuracy. The gaze training group completed the task more quickly and was in general more accurate than the discovery learning group.

TABLE IV
OVERVIEW OF THE EYE-TRACKING STUDIES INVESTIGATING THE IMPACT OF TRAINING ON VIEWING BEHAVIOUR

Source	Wilson et al. [33]	Vine et al. [35]	Krupinski et al. [36]

Modality	Laparoscopic surgery	Laparoscopic surgery	20 breast biopsy whole slide images (10 benign, 10 malignant)
Task	Eye-hand coordination task	Laparoscopic simulator task	Choose 3 areas to zoom in
Partici-pants	30 medical trainees without experience (3 groups: gaze training, movement training, normal feedback	27 naïve participants (2 groups: discovery learning group and gaze learning group)	4 residents tested at 4 four points in time (once a year)
Eye-tracker	ASL mobile	ASL mobile	ASL SU4000
Gaze metrics	Fixation rate, time spent fixating critical locations	Target locking	Fixations, dwell time
Main findings	Group with gaze training made the greatest progress	Gaze learning group improved more strongly (accuracy and time)	Residents became more efficient over time: fewer fixations and revisited locations, longer saccades, less time to make diagnosis

It should be noted that laparoscopic surgery is not the only field where the impact of training was assessed based on eye-tracking. For example, Krupinski et al. [36] studied the impact of training on viewing behaviour in pathology with an ASL SU4000 device. They followed four pathology residents over time during their training, i.e., once a year for four consecutive years. Each time, the residents had to select the top three locations they would like to zoom into in twenty breast core biopsy surgical pathology cases. The fixation positions were recorded, and the dwell time was calculated for each fixation. Statistical analyses showed that the residents became more efficient with training, having fewer fixations generated and fewer locations revisited.

V. A New Eye-tracking Study with Mammograms

The eye-tracking studies reviewed above mainly focused their data analysis on individual fixation locations and durations and used these simple metrics to reveal aspects of human visual behaviour. It would be beneficial for medical imaging to have computational models that can automatically predict human perception. This could help image readers overcome the intrinsic limitations of human perception and reduce diagnostic errors. In the field of computational modelling of visual attention, a topographic representation (i.e., the so-called saliency map) that indicates conspicuousness of scene locations is often used [37]. In a saliency map, the "salient" regions or regions with higher density of fixations designate where the human observers focus their gaze with a higher frequency. In this section, we present a new eye-tracking study, and discuss how to generate ground-truth saliency maps and evaluate to what extent existing computational saliency models can predict human visual attention.

A. Eye-tracking experiment

The source images used in our experiment were acquired from 98 anonymised cases from the University Hospitals KU Leuven, Belgium. They consist of 196 multi-lateral oblique (MLO) views from 98 patients. The original resolution of the mammograms was either 2080×2800 pixels or 2800×3518 pixels. The images (stimuli) were all linearly downscaled to a resolution of 1080×1920 pixels in order to perform an eye-tracking study in a controlled way. The MLO view of the left breast was displayed first to the participants, followed by the MLO view of the corresponding right breast. Each image was displayed for three seconds as to the viewing time in real practice. A 19-inch LCD monitor screen was used and calibrated to the Digital Imaging and Communications in Medicine (DICOM): Grayscale Standard Display Function (GSDF) [38]-[40]. After viewing both images of a case, the participants had to answer the following question: "refer or not refer" by focusing their gaze on one of these two options on the screen. This particular question was asked to simulate the routine breast screening in real practice. Realistically, the suspicious cases would be subjected to further investigation by breast radiologists, but there were none in present database, in line with the screening setting (observers were not informed about this). The aim of the study was to explore their search, not the description of lesions. Fig. 1 illustrates a sequence of the test configuration.

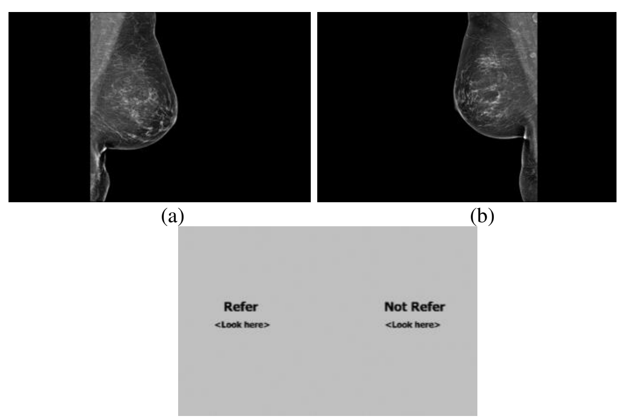


Fig. 1 Illustration of the experimental procedure: (a) represents the MLO view of a left breast, (b) represents the MLO view of the corresponding right breast, and (c) represents the question asked to the participants after viewing (a) and (b).

The experiment was carried out in a mammography reading room in the University Hospitals of the KU Leuven. The venue represented a controlled viewing environment to ensure consistent experimental conditions. The viewing distance was maintained at around 60 cm. The eye movements of the observers were recorded using a non-invasive SensoMotoric

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2851451, IEEE Access

Instrument (SMI) Red-m advanced eye-tracking device at a sampling rate of 250 Hz. Each participant was given written instructions about the procedure prior to the start of the actual experiment, and a training session was conducted to allow the participants to familiarise themselves with the stimuli and the question asked. At the beginning of each session, the eye-tracker was calibrated using a nine-point calibration procedure. Two mammography radiologists, hereafter referred to as R1 and R2, both having more than fifteen years of experience, participated in the eye-tracking experiment. Adding more experts to the study would be highly beneficial, but this was outside the scope of present study. Again, it is worth noting that the goal of this section is to perform a first study to investigate how to create the gaze-based databases, using a computer or having to first expand the experience with more readers, which will be organised if deemed necessary.

B. Gaze duration analysis

Gaze information was extracted directly from the raw eye-tracking data using SMI BeGaze Analysis software, including the coordinates and duration of fixations. A fixation was rigorously defined by SMI's software using the dispersal and duration-based algorithm established in [41] with the minimum fixation duration threshold being set to 100 ms. The average number of fixations per image is 8.9 (with a standard deviation of 1.6) for R1 and 8.1 (with a standard deviation of 2.6) for R2.

As suggested in [8], readers with different degrees of experience can be characterised by their average gaze duration. Fig. 2 represents the mean duration of fixations over all stimuli used in our experiment for each breast radiologist. The average fixation duration is 293.9 ms for R1 and 314.6 ms for R2.

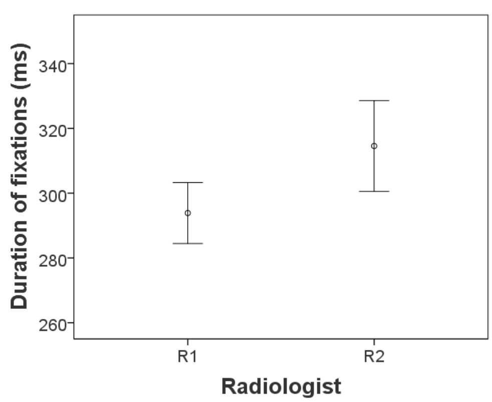


Fig. 2 Illustration of the mean fixation duration for each breast radiologist, averaged over all fixations recorded for all test stimuli. Error bars indicate a 95% confidence interval.

The observed difference was further statistically analysed using hypothesis testing. We first evaluated the assumption of normality of the values of fixation duration, using a Shapiro-Wilk test on R1 and again on R2. The results show that the fixation duration is not normal for R1 and not normal for R2 (i.e., p-value<0.05 in both cases), suggesting that a nonparametric test, the Mann-Whitney U test, for independent samples should be used to reveal the statistical significance between R1 and R2. The results of the Mann-Whitney test show that there is no statistical significant difference in fixation duration between R1 and R2 (i.e., p-value=0.32). The consistency in gaze behaviour between R1 and R2 can be explained by the fact that both observers have substantial experience in mammography screening.

C. Saliency analysis

1) Ground-truth saliency

Eye-tracking data can be also statistically analysed and graphically rendered to explore human visual behaviour. In the area of machine vision, a saliency map is often derived from the recorded fixations. For each stimulus presented to a sample of observers, a saliency map is constructed by accumulating all fixations obtained from eye-tracking with each fixation location giving rise to a greyscale patch that simulates the foveal vision of the human visual system. The activity of the patch is modelled as a Gaussian distribution of which the width approximates the size of the fovea (i.e., 2 degree of visual angle) [42]. Fig. 3 shows the saliency map created from the eye-tracking data for a sample stimulus in our experiment. The saliency map clearly indicates where people look in an image. The brighter the areas, the more salient they are in the given stimulus.

(a) (b)

Fig. 3 Illustration of the saliency map (b) constructed for a sample stimulus (a) used in our experiment.

2) Computational saliency

Eye-tracking is, however, cumbersome and impractical in many circumstances. A more realistic way to integrate gaze information into imaging systems is to use computational saliency. Saliency models, which aim to predict where humans look in images, are available in the literature [43]. These models were developed for different applications, e.g., object detection; however, very little is known about whether these models are directly applicable to medical images and, more specifically, to screening mammography.

To investigate above issues, an evaluation was carried out using three state-of-the-art saliency models, namely Graph Based Visual Saliency model (GBVS), Itti and RARE2012. The GBVS model [44] is a bottom-up visual saliency model composed of two steps including the formation of activation maps and their normalisation to highlight conspicuity. Itti's model [45] was inspired by the neuronal architecture of the primate visual system. Attended locations are selected by a neural network. Finally, RARE2012 [46] selects information based on a multi-scale spatial rarity.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2851451, IEEE Access

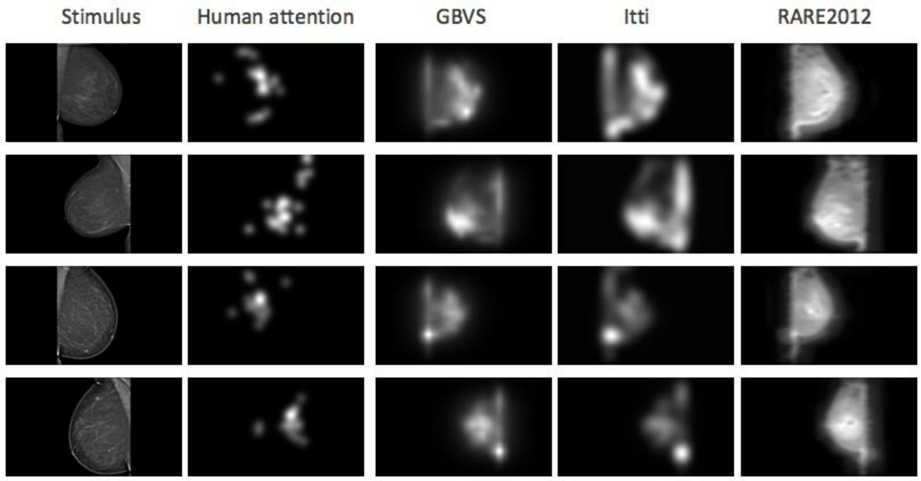


Fig. 4 Illustration of the computational saliency maps generated by three state-of-the-art models (i.e., GBVS, Itti, and RARE2012) for four sample stimuli. Human attention maps resulted from eye-tracking are included in the figure.

Fig. 4 shows the computational saliency maps generated by these three widely used saliency models for four sample stimuli contained in our dataset. It can be seen from the figure that the saliency models do not precisely match with the ground truth (i.e., the "human attention" yielded from fixations of two radiologists R1 and R2).

To quantify the similarity between a saliency model and human fixations, three metrics are commonly used, i.e., the Pearson Correlation Coefficient (CC), the Normalised Scanpath Saliency (NSS), and the Area Under ROC Curve (AUC) [47]. To summarise, when CC is close to -1 or 1, the similarity is high, whereas it is low when CC is close to 0. When NSS>0 or AUC>0.5, the similarity measure is significantly better than chance, and the higher the value is the more similar are the variables.

Fig. 5 illustrates the similarity measures between human and modelled saliency averaged over all stimuli in our database. In general, the CC, NSS and AUC values show a poor correlation with human attention, e.g., all CC values are less than 0.6. This suggests that a more accurate saliency modelling is needed to better predict the viewing behaviour of radiologists when evaluating mammograms.

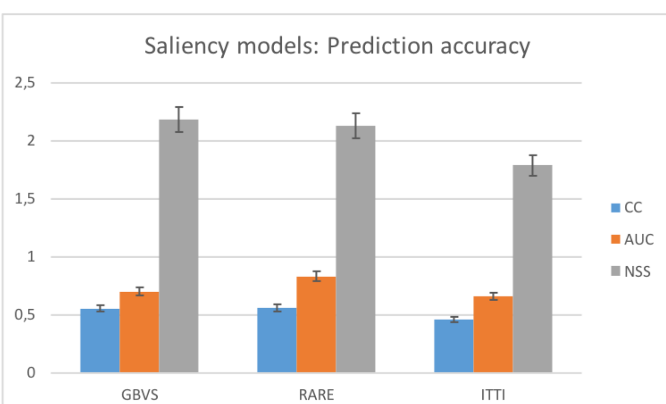


Fig 5. Illustration of the similarity measures between human and modelled saliency averaged over the 194 stimuli using the CC, NSS, and AUC metrics. Error bars indicate a 95% confidence interval.

VI. Conclusions

In this paper, we have reviewed state-of-the-art eye-tracking studies in the area of medical imaging. We have evaluated their motivations, methodologies for data collection and analysis, and significant findings. There is evidence of the importance of integrating aspects of human visual attention to imaging systems, so that advanced computational tools can be of benefit to readers and aid in the interpretation of medical images. To add value to the survey, we present a new eye-tracking study, where a large-scale database of mammograms was assessed by two expert radiologists. Based on the resulting eye-tracking data, we aim to investigate the plausibility of predicting human visual attention by use of computational models. It is clear that computer-generated saliency maps cannot sufficiently predict the human gaze behaviour yet, and therefore further improvements are needed before they can be used to automatically generate large databases for gaze-based training purposes.

References

[1] T. Ohno, "One-point calibration gaze tracking method", *Proceedings on the 2006 symposium on Eye tracking research and applications, San Diego, California*, pp. 34-34, 2006.

[2] C. Connor, H. Egeth, and S. Yantis, "Visual attention: bottom-up versus top-down", *Current Biology*, vol. 14, pp. 850-852, 2004.

[3] W. Zhang, and H. Liu, "Toward a Reliable Collection of Eye-Tracking Data for Image Quality Research: Challenges, Solutions, and Applications," *IEEE Transactions on Image Processing*, vol. 26, pp. 2424-2437, 2017.

[4] G. Buswell, How people look at pictures: A study of the psychology of perception in art, The University of Chicago Press, 1935.

[5] A. Borji, and L. Itti, "State-of-the-art in visual attention modelling", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, 2013.

[6] L. Lévêque, W. Zhang, P. Parker, and H. Liu, "The impact of specialty settings on the perceived quality of medical ultrasound video", *IEEE Access*, vol. 5, pp. 16998-17005, 2017.

[7] A. Brady, "Error and discrepancy in radiology: inevitable or avoidable?", *Insights Imaging*, vol. 8, pp. 171-182, 2017.

[8] E. Krupinski, "Visual scanning patterns of radiologists searching mammograms", *Academic Radiology*, vol.3, pp.137-144, 1996.

[9] E. Krupinski, "Current perspectives in medical image perception", *Attention, Perception, & Psychophysics*, vol. 72, pp. 1205-1217, 2010.

[10] D. Carmody, C. Nodine, and H. Kundel, "Finding lung nodules with and without comparative visual scanning", *Perception & Psychophysics*, vol. 29, pp. 594-598, 1981.

[11] D. Beard, R. Johnston, O. Toki, and C. Wilcox, "A Study of Radiologists Viewing Multiple Computed Tomography Examinations Using an Eyetracking Device", *Journal of Digital Imaging*, vol. 4, pp. 230-237, 1990.

[12] K. Suwa, A. Furukawa, T. Mastumoto, and T. Yosue, "Analyzing the eye movement of dentists during their reading of CT images", *Odontology*, vol. 89, pp. 54-61, 2001.

[13] H. Kundel, C. Nodine, E. Krupinski and C. Mello-Thoms, "Using Gaze-tracking Data and Mixture Distribution Analysis to Support a Holistic Model for the Detection of Cancers on Mammograms", *Academic Radiology*, vol. 15, pp. 881-886, 2008.

[14] S. Voisin, F. Pinto, S. Xu, G. Morin-Ducote, K. Hudson, and G. Tourassi, "Investigating the Association of Eye Gaze Pattern and Diagnostic Error in Mammography", *Proceedings of Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, vol. 8673, 2013.

[15] C. Almansa, M. Shahid, M. Heckman, S. Preissler, and M. Wallace, "Association Between Visual Gaze Patterns and Adenoma Detection Rate During Colonoscopy: A Preliminary Investigation", *The American Journal of Gastroenterology*, vol. 106, pp. 1070-1074, 2011.

[16] T. Drew, M. Le-Hoa Vo, A. Olwal, F. Jacobson, S. Seltzer, and J. Wolfe, "Scanners and drillers: Characterizing expert visual search through volumetric images", *Journal of Vision*, vol. 13, pp. 1-13, 2013.

[17] C. Nodine, C. Mello-Thoms, H. Kundel, and S. Weinstein, "Time Course of Perception and Decision Making During Mammographic Interpretation", *American Journal of Roentgenology*, vol. 179, pp. 917-923, 2002.

[18] G. Tourassi, S. Voisin, V. Paquit, and E. Krupinski, "Investigating the link between radiologists' gaze, diagnostic decision, and image content", *Journal of the American Medical Informatics Association*, vol. 20, pp. 1067-1075, 2013.

[19] L. Cooper, A. Gale, I. Darker, A. Toms, and J. Saada, "Radiology image perception and observer performance: how does expertise and clinical information alter interpretation? Stroke detection explored through eye-tracking.", *Proceedings of Medical Imaging: Image Perception, Observer Performance, and Technology Assessment*, vol. 7263, 2009.

[20] H. Matsumoto, Y. Terao, A. Yugeta, H. Fukuda, M. Emoto, T. Furubayashi, T. Okano, R. Hanajima, and Y. Ugawa, "Where Do Neurologists Look When Viewing Brain CT Images? An Eye-Tracking Study Involving Stroke Cases", *PLoS One*, vol. 6, 2011.

[21] R. Bertram, L. Helle, J. Kaakinen, and E. Svedstrom, "The Effect of Expertise on Eye Movement Behaviour in Medical Image Perception", *PLoS One*, vol. 8, 2013.

[22] R. Bertram, J. Kaakinen, F. Bensch, L. Helle, E. Lantto, P. Niemi, and N. Lundbom, "Eye Movements of Radiologists Reflect Expertise in CT Study Interpretation: A Potential Tool to Measure Resident Development", *Radiology*, vol. 281, 2016.

[23] S. Mallett, P. Philips, T. Fanshawe, E. Helbren, D. Boone, A. Gale, S. Taylor, D. Manning, D. Altman, and S. Halligan, "Tracking Eye Gaze during Interpretation of Endoluminal Three-dimensional CT Colonography: Visual Perception of Experienced and Inexperienced Readers", *Radiology*, vol. 273, 2014.

[24] D. Manning, S. Ethell, T. Donovan, and T. Crawford, "How do radiologists do it? The influence of experience and training on searching for chest nodules", *Radiography*, vol. 12, pp. 134-142, 2006.

[25] J. Leong, M. Nicolaou, R. Emery, A. Darzi, and G. Yang, "Visual search behaviour in skeletal radiographs: a cross-specialty study", *Clinical Radiology*, vol. 62, pp. 1069-1077, 2007.

[26] P. Vaidyanathan, J. Pelz, C. Alm, P. Shi, and A. Haake, "Recurrence Quantification Analysis Reveals Eye-Movement Behavior Differences between Experts and Novices", *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 303-306, 2014.

[27] D. Turgeon, and E. Lam, "Influence of Experience and Training on Dental Students' Examination Performance Regarding Panoramic Images", *Journal of Dental Education*, vol. 80, pp. 156-164, 2016.

[28] B. Law, M. Atkins, A. Kirkpatrick, A. Lomax, and C. Mackenzie, "Eye Gaze Patterns Differentiate Novice and Experts in a Virtual Laparoscopic Surgery Training Environment", *Proceedings of the Symposium on Eye tracking research and applications, San Antonio, USA*, pp. 41-48, 2004.

[29] E. Kocak, J. Ober, N. Berme, and W. Melvin, "Eye Motion Parameters Correlate with Level of Experience in Video-Assisted Surgery: Objective Testing of Three Tasks", *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol. 15, 2005.

[30] N. Ahmidi, G. Hager, L. Ishii, G. Fichtinger, G. Gallia, and M. Ishii, "Surgical Task and Skill Classification from Eye Tracking and Tool Motion in Minimally Invasive Surgery", *International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 13, pp. 295-302, 2010.

[31] L. Richstone, M. Schwartz, C. Seideman, J. Cadeddu, S. Marshall, and L. Kavoussi, "Eye Metrics as an Objective Assessment of Surgical Skill", *Annals of Surgery*, vol. 252, pp.177-182, 2010.

[32] R. Khan, G. Tien, M. Atkins, B. Zheng, O. Panton, A. Meneghetti, "Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation?", *Surgical Endoscopy*, vol. 26, pp. 3536-3540, 2012.

[33] M. Wilson, S. Vine, E. Bright, R. Masters, D. Defriend, and J. McGrath, "Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study", *Surgical Endoscopy*, vol. 25, pp. 3731-3739, 2011.

[34] M. Land, "Vision, eye movements, and natural behavior", *Visual Neuroscience*, vol. 26, pp. 51-62, 2009.

[35] S. Vine, R. Masters, J. McGrath, E. Bright, and M. Wilson, "Cheating experience: Guiding novices to adopt the gaze strategies of experts expedites the learning of technical laparoscopic skills", *Surgery*, vol. 152, pp. 32-40, 2012.

[36] E. Krupinski, A. Graham, and R. Weinstein, "Characterizing the development of visual search expertise in pathology residents viewing whole slide images", *Human Pathology*, vol. 44, pp. 357-364, 2013.

[37] H. Liu, and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data", *IEEE Transactions Circuits and Systems for Video Technologies*, vol. 21, pp. 971-982, 2011.

[38] E. Samei, et al., "Assessment of display performance for medical imaging systems", *Medical Physics*, vol. 32, pp. 1205-1225, 2005.

[39] B. Hemminger, R. Johnston, J. Rolland, and K. Muller, "Introduction to perceptual linearization of video display systems for medical image presentation", Journal of Digital Imaging, vol. 8, pp. 21-34, 1995.

[40] National Electrical Manufacturers Association, *Digital Imaging and Communications in Medicine (DICOM): Part 14 Greyscale Standard Display Function*, [online] last accessed in Feb. 2018, http://medical.nema.org/dicom/2004/04_14PU.pdf

[41] D. Salvucci, and J. Goldberg, "Identifying fixations and saccades in eye-tracking protocols", *Proceedings of the Eye Tracking Research and Applications Symposium, Tampa, USA*, pp. 71-78, 2000.

[42] C. Privitera, and L. Stark, "Algorithms for defining visual regions-of-interest: comparison with eye fixations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 970-982, 2000.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2851451, IEEE Access

[43] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations", *MIT Technical Report*, 2012.

[44] J. Harel, C. Koch, and P. Perona, "Graph-Based Visual Saliency", *Proceedings of Advances of Neural Information Processing Systems, Pasadena, Canada*, pp. 545-552, 2006.

[45] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1254-1259, 1998.

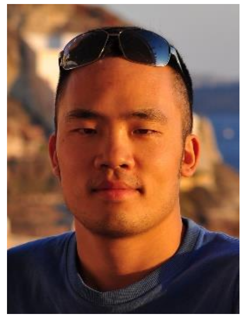
[46] N. Riche, M. Mancas, M. Duvinage, M. Mibulumukini, B. Gosselin, and T. Dutoit, "RARE2012 : A multi-scale rarity based saliency detection with its comparative statistical analysis", Image Communication, vol. 28, pp. 642-658, 2013.

[47] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?", arXiv:1611.09571, 2016.

Lucie Lévêque received the B.Eng. and M.Eng. degrees in cognitics from the Ecole Nationale Supérieure de Cognitique, Bordeaux, France, in 2013 and 2015, respectively, and the M.Sc. degree in biomedical imaging from the University of Angers, Angers, France, in 2015. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering at Cardiff University, Cardiff, U.K. Her research interests include visual media quality assessment, and medical image perception.

Hilde Bosmans received the Ph.D. degree from the faculty of Medicine of the Katholieke Universiteit Leuven, Leuven, Belgium, in 1992. She is currently a Professor at KU Leuven, Leuven, Belgium. Her research interests include medical radiation physics, patient dosimetry, image perception, and image quality assessment. KU Leuven and Cardiff University have bilateral research contacts.

Lesley Cockmartin received the Ph.D. degree from the Department of Imaging and Pathology of the Katholieke Universiteit Leuven, Leuven, Belgium, in 2015. She is currently a medical physicist at the University Hospitals of KU Leuven, Leuven, Belgium. Her research interests include digital mammography, medical physics, and image quality assessment.



Hantao Liu (S'07-M'11) received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Assistant Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. Prof. Liu is currently serving for IEEE MMTC as the Chair of the Interest Group on Quality of Experience for Multimedia Communications, and he is an Associate Editor of the IEEE Transactions on Human-Machine Systems and of the IEEE Transactions on Multimedia. His research interests include visual media quality assessment, visual attention modeling and applications, visual scene understanding, and medical image perception.